A Low Rank Model based Improved Eye Detection under Spectacles

M Zefree Lazarus
Dept. of Electrical Engineering
National Institute of Technology
Rourkela, India 769008
Email: zefree.lazarus@gmail.com

Supratim Gupta
Dept. of Electrical Engineering
National Institute of Technology
Rourkela, India 769008
Email: sgupta.iitkgp@gmail.com

***Abstract*—Eye detection is a primary step in many applications such as face recognition, iris recognition, driver fatigue detection, gaze tracking etc. Occlusion by spectacles, glare and secondary image formations deteriorate its performance. In this paper, we formulate the glare/reflection removal as a classification problem and employ Low rank decomposition technique to overcome these challenges. We provide an in-depth analysis by comparing various low rank decomposition formulations and propose a simple preprocessing step to improve the detection accuracy. Experimentation on CASIA NIR-VIS 2.0 facial database validates the proposed preprocessing method.**

***Keywords*—Eye detection, Low rank model, Spectacle reflection, Glare removal.**

I. Introduction

Detecting eyes accurately is the first step in real time applications like face detection, eye gaze estimation and many other human computer interactive (HCI) systems. Occlusion due to spectacles, glare and secondary image formations on eyes put challenges in its detection.

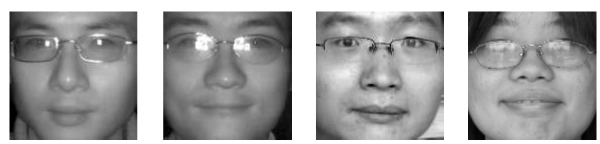
A. The Spectacle problem

Song et al. [1] have divided eye detection algorithms in five categories like: Feature based algorithms [2], Appearance based algorithms [3], Shape based algorithms [4], Sparse representation based algorithms [5] and Hybrid algorithms [6]. However these algorithms performs poorly when users wear spectacles and the ambient illumination is variable [7].

Though several studies on the spectacle detection and removal have been reported [8–11], the study focusing on secondary reflection and glare removal is very limited. Literature suggests that few methods employ image inpainting and supervised learning techniques to mitigate the spectacle problem. Inpainting utilizes neighbouring pixel values to create a seamless reconstructed image. As glare and reflection occlude the underlying eye feature information, it is difficult to reconstruct a reflectionless image [12],[13]. The representational power of supervised learning methods totally depend upon the exhaustiveness of its training set [11]. For example, it is hard to reconstruct a spectacle-less image from a training set of images containing spectacles only. A few of the problems cited in the literature are given in Table I.

The problems pertaining to the usage of spectacles can be summarized as:

- *Occlusion*: In applications like eye feature detection, facial landmarking [21], important eye features like eyelids, eyebrows [12], [22] could be occluded by the spectacle frame (see Fig.1c) resulting in performance inaccuracies.
- *Illumination changes*: Appearance of eye features changes drastically with respect to illumination. These uncontrolled illumination changes causes glare and secondary reflection on spectacles. Occlusion due to glare (see Fig.1a,1b), appearance change due to the secondary reflection (see Fig.1d) cause great challenges to many eye detection algorithms [10], [23] .



(a) (b) (c) (d)

Fig. 1: Sample images illustrating the Spectacle problem. (a),(b) Glare; (c) Occlusion; (d) Secondary reflection

Under these unknown and varying conditions, it is very challenging to nullify the reflection effect on the spectacles. These inconsistencies in specular reflection removal are yet to be thoroughly investigated and there is scope of improvement. In this paper we present a preprocessing step which mitigates the ill-effects of spectacles to improve detection of eyes under it.

The rest of the article is organized as follows: The concept and mathematical framework of the proposed preprocessing step are discussed in Section II. Experimental evaluation of various Low rank decomposition strategies are presented in Section III. The experimental results of eye detection under spectacles are provided and discussed in Section IV. Conclusion of this paper is given in Section V.

II. Problem formulation

A. The concept

We model the reflection due to spectacles as sparse error (S) and separate the face as low rank component (L) from the captured image (M) as shown in Fig.2.

TABLE I: Literature stating the spectacle problem

Literature	Application of the citation	Problem reported
[1],[14]	Eye detection	Locating eyes under occlusion is largely problematic. Poor performance under spectacles.
[15],[16]	Eye detection & Tracking	With the secondary reflections on the spectacles, detection and tracking produces erroneous results.
[17],[4]	Eye center detection	Strong reflections on the spectacles result in inaccurate estimations of eye center.
[6],[18]	Iris center & Eye corners detection	Secondary reflections occlude iris resulting in poor performance.
[19],[20]	Pupil Center, Eye corner & Eyelids detection	Spectacles occlude the eye boundary detection.
[21]	Facial landmarking	Visual obstructions like spectacles reduce the performance of local feature detectors.

(a) (b) (c)

Fig. 2: The separation problem. (a) Corrupted input image; (b) Low rank component; (c) Sparse component

Thus we can denote the problem mathematically as

$$M = L + S \tag{1}$$

where M is the corrupted data matrix; L is the Low rank component and S is the Sparse component.

Literature indicate that the facial images under varying illumination conditions lie on a low dimensional manifold [24] and low rank matrix can model them effectively [25]. Glare or secondary reflection may be interpreted as the error signal. Depending on the illumination type, the error is either concentrated (as in case of glare) or distributed (as in case of secondary reflection). However in both cases, only few eye region pixels are corrupted compared to the total face image pixels. Therefore, the error signal is formulated as the sparse component, as shown in Eq.1.

B. Mathematical Framework

Researchers have developed a number of approaches to solve Eq.1. All these approaches lead to the following formulation similar to Principal Component Pursuit (PCP) [25].

$$\begin{aligned} &Minimize\ ||L||_{\star} + \lambda ||S||_1 \\ &Subject\ to\ L + S = M \end{aligned} \tag{2}$$

where, $||.||_1$ and $||.||_{\star}$ denote the l_1-norm and the nuclear norm respectively, and $\lambda > 0$ is an arbitrary weighting parameter..

In most of the real-time situations, the details of the location, nature and intensity of the corruption is unknown. Even under these circumstances, the decomposition formulation still holds good and the desirable uncorrupted data can be extracted. Another advantage of this formulation is that it requires minimum tuning parameters.

Recently, new implicit and explicit problem formulations for low rank decomposition have been developed. The core idea of the decomposition remains the same in all these approaches, however they differ in the way the loss function and optimization problem is modelled. The mathematical models can be broadly divided into two categories:

i. Matrix-based algorithms
   - Robust Principal Component Analysis (RPCA)
   - Subspace Tracking (ST)
   - Matrix Completion (MC)
   - Three-Term Decomposition (TTD)
   - Low-Rank Representation (LRR)
   - Non-negative Matrix Factorization (NMF)

ii. Tensor-based algorithms
   - Non-negative Tensor Factorization (NTF)
   - Tensor Decomposition (TD)

Six Matrix and two Tensor based algorithms have been chosen for conducting different experiments. These eight methods were considered mainly because they are prevalent and successfully implemented in many Low rank decomposition applications. Moreover, the study of these methods will give us an exhaustive insight on how each approach differs from the other and to identify the one which suits the best for our application. The details and formulation of these algorithms for Low rank decomposition can be found in [26].

III. Experimental evaluation

In this work, we are analyzing the best possible mathematical formulation to improve the performance of eye detection method under the spectacles problem.

A. The database

In literature, most of the facial image databases are created under well controlled laboratory conditions. These databases are mostly used for development of face recognition or eye feature extraction algorithms. There are very few databases created in uncontrolled environments with pairs of corresponding facial images with and without spectacles. We found that CASIA NIR-VIS 2.0 Face Database [27] provides the desired requirements. It provides linearly correlated heterogeneous (both NIR and Visible) facial images of size 128×128, taken from 725 subjects. The details of the database is given in Table II

For conducting experiments, we have segregated the images of each illumination type into two categories.

- NIR_Db_1: consists of 921 facial images with spectacles only.
- NIR_Db_2: consists of 292 images, of which 146 are facial images with spectacles and the rest are images without spectacles.

TABLE II: Details of CASIA NIR-VIS 2.0 database

Visible Light		NIR illumination	
No. of images with spectacles	No. of images without spectacles	No. of images with spectacles	No. of images without spectacles
1248	222	1067	146

- VIS_Db_1: consists of 1024 facial images with spectacles only.
- VIS_Db_2: consists of 444 images, of which 222 are facial images with spectacles and the rest are images without spectacles.

We conducted the experiments on these four sub-categorized databases to validate our analysis and inference.

B. *Experiment* 1

Many mathematical formulations and algorithms as mentioned in Section II-B, are available in the literature for the implementation of Low rank decomposition. The aim of this experiment is to identify the algorithm that suits best for this particular application. Initially, images from the NIR_Db_1 dataset are vectorized and linearly stacked as a corrupted data matrix. One algorithm of each mathematical model is considered for decomposition. Example output image of each model is shown in Fig 3.

It is quite evident from the visual inspection of Fig 3 that the RPCA approach retains more low rank information (facial features) than the remaining formulations. So further experimentation is carried to find the best available RPCA implementation algorithm.

C. *Experiment* 2

Several implementation variations of RPCA can be found in the literature [28]. These algorithms are quite similar to each other but they differ in the way the SVD and convex relaxation is implemented. In this paper, we have considered the following 8 algorithms for experimentation:

- Principal Component Pursuit (PCP)
- Exact Augmented Lagrange Multiplier (EALM)
- Inexact Augmented Lagrange Multiplier (IALM)
- IALM with block Lanczos with warm start (IALM_BLWS)
- Accelerated Proximal Gradient (APG)
- Partial Accelerated Proximal Gradient (APG_PARTIAL)
- Alternating Direction Method (LSADM)
- Non-Smooth Augmented Lagrangian v1 (NSA1)

Images from the NIR_Db_1 and NIR_Db_2 datasets are vectorized and linearly stacked as two corrupted input data matrices respectively. The above mentioned Low rank decomposition implementation strategies are applied on the input matrices and the results are shown in Tables III and IV.

D. *Experiment* 3

The effects of glare and reflections are not the same in the case of visible light and NIR illuminated images. The chances of glare formation is more in case of NIR illumination and the chances of secondary reflection formation is more in case of Visible light illumination [29]. To verify the performance of the algorithms under visible light illumination, we repeated the experiments as in Section III-B and III-C with VIS_Db_1 and VIS_Db_2 databases.

E. *Experiment* 4

To validate the performance of the proposed preprocessing step, an eye-pair detection experiment was conducted on the above mentioned four datasets. AdaBoost-based eye-pair detection algorithm [30] was employed for the experimentation. Sample results of the eye pair detection algorithm conducted on NIR_Db_1 and VIS_Db_1 datasets is shown in Fig.5.

All the experiments were conducted in batch mode with MATLAB $R2015a$ on a PC with Intel (R) core (TM) $i7-4770$ CPU and 4.00 GB RAM.

IV. Results and discussion

The effectiveness of the algorithms (Section III-C, III-D) is compared based on rank minimization capability and the time taken for execution. Rank minimization capability (L-rank) emphasises the feature retaining property of the algorithm. It is the lowest possible rank that the decomposition can attain without degrading the original feature content beyond a permissible error. The computational cost of the RPCA algorithms mainly depends on SVD implementation.

TABLE III: Experimental results conducted on NIR_Db_1 dataset. M-rank of input matrix, L-rank of low rank matrix.

S.No	Algorithm	M-rank	L-rank	Time in Sec.
1	PCP	921	464	391.3956
2	EALM	921	634	877.0114
3	IALM	921	517	63.8970
4	IALM_BLWS	921	487	934.2517
5	APG	921	547	376.5881
6	APG_PARTIAL	921	580	283.5623
7	LSADM	921	921	237.4707
8	NSA1	921	572	654.4221

TABLE IV: Experimental results conducted on NIR_Db_2 dataset. M-rank of input matrix, L-rank of low rank matrix.

S.No	Algorithm	M-rank	L-rank	Time in Sec.
1	PCP	292	141	53.6680
2	EALM	292	182	162.0471
3	IALM	292	151	12.5318
4	IALM_BLWS	292	144	107.5765
5	APG	292	154	84.4015
6	APG_PARTIAL	292	159	65.174
7	LSADM	292	290	35.0362
8	NSA1	292	175	176.7217

Tables III and IV indicates that PCP and LSADM algorithms gives the least and highest possible rank decompositions. PCP algorithm does not retain most of the features (low L-rank value) where as LSADM algorithm retains back the original corrupted data. In both the cases eye detection may not be accurate due to lack of proper eye feature information after the decomposition.

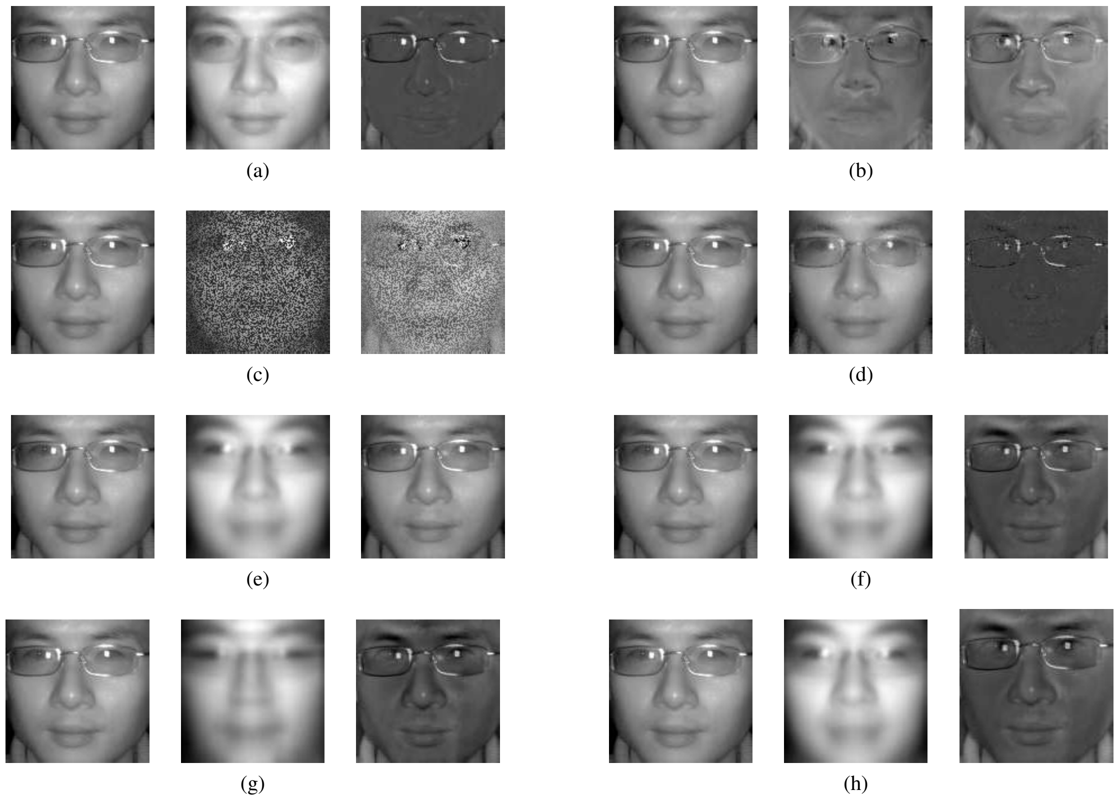


Fig. 3: Example output images of Low rank decomposition by (a) RPCA (b) ST (c) MC (d) TTD (e) LRR (f) NMF (g) NTF (h) TD algorithms. Each sub-plot contains the input, low rank and sparse component images respectively.

For real-time applications like eye detection, computation time is prioritized than the localization accuracy of the system. Among the 8 algorithms, IALM has the least average running time of 0.063 seconds per image. Although EALM gives better L-rank value, it's computation time is approximately 13 times more than the IALM algorithm. This significant reduction in time is because, the IALM algorithm employs partial SVD implementation approach. Due to its minimum time complexity and adequate rank minimization capability, we conclude that IALM algorithm suits best for the proposed preprocessing step. Fig.4 shows the sample results of Low rank decomposition using IALM algorithm.

The results (from Section III-E) of the eye-pair detection before and after applying the pre-processing step can be seen in Table V.

TABLE V: Number of successful eye-pair detections before and after the Low rank decomposition (IALM) preprocessing step.

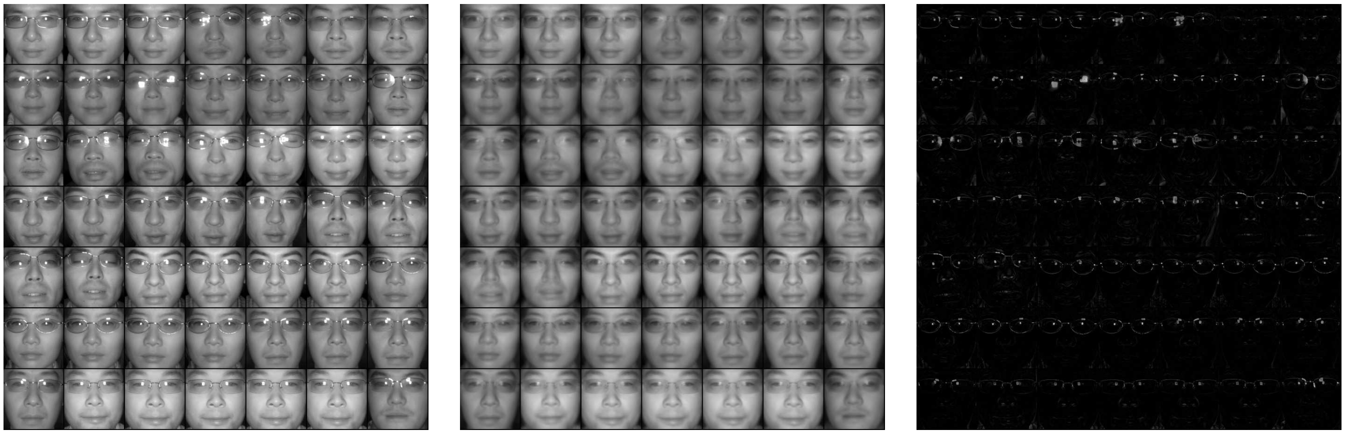
Database	Before preprocessing		After preprocessing	
	No. of unsuccessful detections	No. of successful detections	No. of unsuccessful detections	No. of successful detections
NIR_Db_1	692	229	362	559
NIR_Db_2	111	181	35	257
VIS_Db_1	178	846	77	947
VIS_Db_2	14	430	12	432

Table V indicate that before preprocessing, the number of eye-pair detections are less due to the influence of specular and glare reflections. Particularly, NIR illuminated facial images have more unsuccessful detections than the VIS illuminated images. After employing the preprocessing step, the number of detections have increased significantly. It can also be observed that the rate of increase in detections is more significant in case of interclass datasets as compared to the intraclass datasets. Many of the unsuccessful detections are due to the presence of spectacle frame residue even after applying the low rank decomposition (Fig. 5b). Another reason for the unsuccessful detection is that, majority of the input eye features are heavily corrupted, making a seamless eye recovery impossible (Fig. 5a). Even though the processed low rank images (Fig.4) appear to be slightly blurred, the quality of the image features is enough for eye region detection.

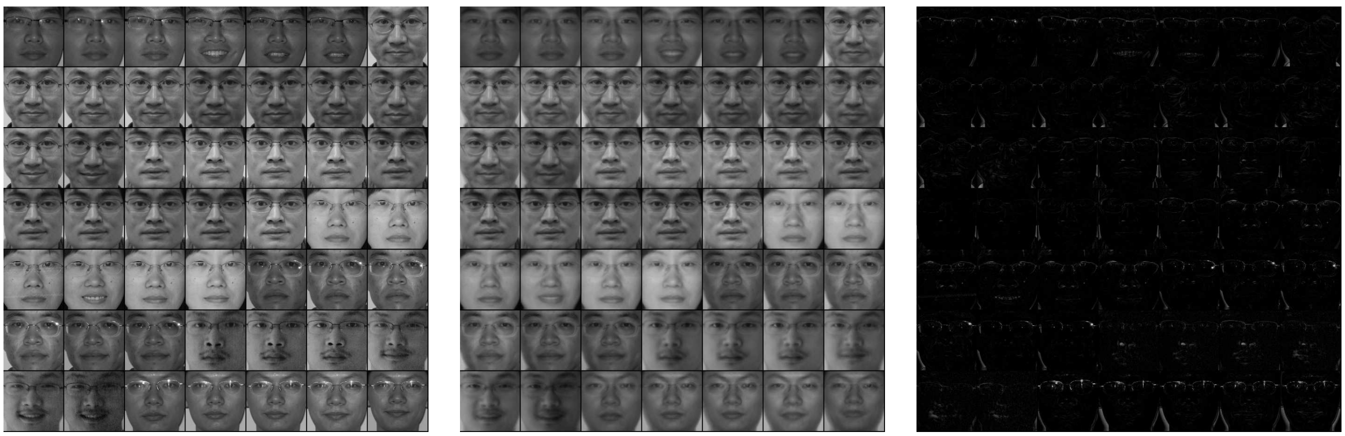
The overall experimental results demonstrates that the Low rank decomposition based preprocessing step is effective and efficient.

V. Conclusion

Spectacles cause hindrance to the eye detection application. We alleviate the problem by formulating it as sparse component and removed it from the low rank part of the face. Experimentation with different types of Low rank decomposition formulations were carried on CASIA NIR-VIS 2.0 Face



(a)



(b)

Fig. 4: IALM based Low rank decomposition outputs of 49 sample images conducted on (a) NIR_Db_1 (b) VIS_Db_1 database. Each sub-plot contains original, low rank and sparse component images respectively.

Database. Both qualitative and quantitative results indicate that IALM algorithm suits best for the eye detection application. After applying the Low rank decomposition, the number of successful eye pair detections increased significantly without increasing the computational complexity of the eye detection algorithm considerably.

Although Low rank decomposition provides a reliable solution, it is a batch processing approach which needs stacking of a series of images. The assumption of face alignment might not be true in most of the cases. The authors are engaged to improve its performance.

ACKNOWLEDGMENT

This work was partially supported by the Board of Research in Nuclear Sciences (BRNS), Government of India, under the grant number 34/14/08/2016.

REFERENCES

[1] F. Song, X. Tan, S. Chen, and Z.-H. Zhou, "A literature survey on robust and efficient eye localization in real-life scenarios," *Pattern Recognition*, vol. 46, no. 12, pp. 3157–3173, 2013.

[2] S. Kim, S.-T. Chung, S. Jung, D. Oh, J. Kim, and S. Cho, "Multi-scale gabor feature based eye localization," *World Academy of Science, Engineering and Technology*, vol. 21, pp. 483–487, 2007.

[3] M. Everingham and A. Zisserman, "Regression and classification approaches to eye localization in face images," in *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*. IEEE, 2006, pp. 441–446.

[4] R. Valenti and T. Gevers, "Accurate eye center location through invariant isocentric patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 9, pp. 1785–1798, 2012.

[5] C. Meng and T. Zhang, "Human eye detection via sparse representation," in *Proceedings of the 2015 International Conference on Communications, Signal Processing, and Systems*. Springer, 2016, pp. 693–700.

[6] A. Villanueva, V. Ponz, L. Sesma-Sanchez, M. Ariz, S. Porta, and R. Cabeza, "Hybrid method based on topography for robust detection of iris center and eye corners," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 9, no. 4, p. 25, 2013.

[7] Y. Wu and Q. Ji, "Learning the deep features for eye detection in uncontrolled conditions." in *ICPR*, 2014, pp. 455–459.

[8] W.-C. Cheng, H.-C. Liao, M.-H. Pan, and C.-C. Chen, "A fatigue detection system with eyeglasses removal," in *Advanced*

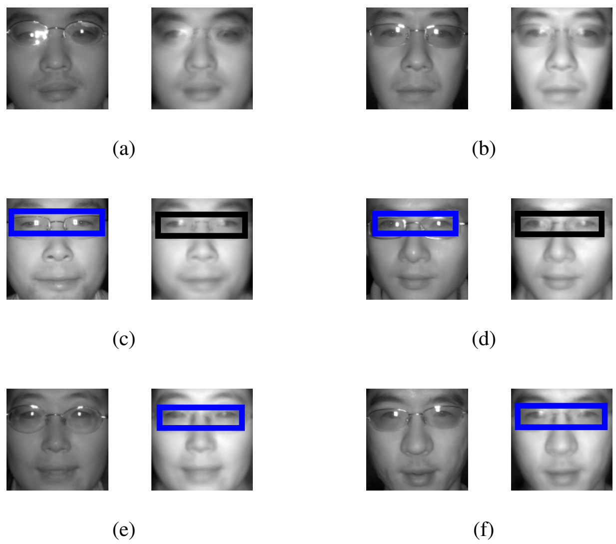


Fig. 5: Example results of (a) unsuccessful, (c,e) successful eye-pair detection on NIR_Db_1 database; (b) unsuccessful, (d,f) successful eye-pair detection on VIS_Db_1 database. Each sub-plot contains eye-pair detection output image before and after the preprocessing.

Communication Technology (ICACT), 2013 15th International Conference on. IEEE, 2013, pp. 331–335.

[9] Y.-K. Wang, J.-H. Jang, L.-W. Tsai, and K.-C. Fan, "Improvement of face recognition by eyeglass removal," in *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2010 Sixth International Conference on*. IEEE, 2010, pp. 228–231.

[10] J.-S. Park, Y. H. Oh, S. C. Ahn, and S.-W. Lee, "Glasses removal from facial image using recursive pca reconstruction," in *Audio- and Video-Based Biometric Person Authentication*. Springer, 2003, pp. 369–376.

[11] C. Du and G. Su, "Eyeglasses removal from facial images," *Pattern Recognition Letters*, vol. 26, no. 14, pp. 2215–2220, 2005.

[12] X. Jia and J. Guo, "Eyeglasses removal from facial image based on phase congruency," in *Image and Signal Processing (CISP), 2010 3rd International Congress on*, vol. 4. IEEE, 2010, pp. 1859–1862.

[13] C. Wu, C. Liu, H.-Y. Shum, Y.-Q. Xy, and Z. Zhang, "Automatic eyeglasses removal from face images," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 3, pp. 322–336, 2004.

[14] G.-C. Feng and P. C. Yuen, "Variance projection function and its application to eye detection for human face recognition," *Pattern Recognition Letters*, vol. 19, no. 9, pp. 899–906, 1998.

[15] Y.-l. Tian, T. Kanade, and J. F. Cohn, "Dual-state parametric eye tracking," in *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*. IEEE, 2000, pp. 110–115.

[16] Z. Pang, C. Wei, D. Teng, D. Chen, and H. Tan, "Robust eye center localization through face alignment and invariant isocentric patterns," *PloS one*, vol. 10, no. 10, p. e0139098, 2015.

[17] F. Timm and E. Barth, "Accurate eye centre localisation by means of gradients." *VISAPP*, vol. 11, pp. 125–130, 2011.

[18] Z. Zhu and Q. Ji, "Robust real-time eye detection and tracking under variable lighting conditions and various face orientations," *Computer Vision and Image Understanding*, vol. 98, no. 1, pp. 124–154, 2005.

[19] Z. Zheng, J. Yang, and L. Yang, "A robust method for eye features extraction on color image," *Pattern Recognition Letters*, vol. 26, no. 14, pp. 2252–2261, 2005.

[20] T. Moriyama, T. Kanade, J. Xiao, and J. F. Cohn, "Meticulously detailed eye region model and its application to analysis of facial images," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 5, pp. 738–752, 2006.

[21] P. N. Belhumeur, D. W. Jacobs, D. J. Kriegman, and N. Kumar, "Localizing parts of faces using a consensus of exemplars," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 12, pp. 2930–2940, 2013.

[22] D. Liu, L. Shen, Y. Yin, and X. Li, "How to recognize facial images with spectacles," in *Intelligent Control and Automation, 2006. WCICA 2006. The Sixth World Congress on*, vol. 2. IEEE, 2006, pp. 10 153–10 156.

[23] J. Gómez-Poveda and E. Gaudioso, "Evaluation of temporal stability of eye tracking algorithms using webcams," *Expert Systems with Applications*, vol. 64, pp. 69–83, 2016.

[24] R. Basri and D. W. Jacobs, "Lambertian reflectance and linear subspaces," *IEEE transactions on pattern analysis and machine intelligence*, vol. 25, no. 2, pp. 218–233, 2003.

[25] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM (JACM)*, vol. 58, no. 3, p. 11, 2011.

[26] T. Bouwmans, A. Sobral, S. Javed, S. K. Jung, and E.-H. Zahzah, "Decomposition into low-rank plus additive matrices for background/foreground separation: A review for a comparative evaluation with a large-scale dataset," *arXiv preprint arXiv:1511.01245*, 2015.

[27] S. Li, D. Yi, Z. Lei, and S. Liao, "The casia nir-vis 2.0 face database," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2013, pp. 348–353.

[28] A. Sobral, T. Bouwmans, and E.-h. Zahzah, "Lrslibrary: Low-rank and sparse tools for background modeling and subtraction in videos," in *Robust Low-Rank and Sparse Matrix Decomposition: Applications in Image and Video Processing*. CRC Press, Taylor and Francis Group.

[29] S. Z. Li, S. R. Chu, S. Liao, and L. Zhang, "Illumination invariant face recognition using near-infrared images," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 4, pp. 627–639, 2007.

[30] M. Castrillón, O. Déniz, C. Guerra, and M. Hernández, "Encara2: Real-time detection of multiple faces at different resolutions in video streams," *Journal of Visual Communication and Image Representation*, vol. 18, no. 2, pp. 130–140, 2007.